

DOCUMENT RESUME

ED 165 712

IR 006 275

AUTHOR Weaver, Tamar A.
TITLE On-Line Data in CBE.
PUB DATE Mar 78
CONTRACT DAHC-15-73-C-0077; NSF-C-723
NOTE 16p.; Paper presented at the Annual Meeting of the Association for the Development of Computer Based Instructional Systems (Dallas, Texas, March 1-4, 1978); For related document, see ED 160 072

EDRS PRICE MF-\$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Computer Assisted Instruction; *Computer Managed Instruction; Educational Research; Instructional Design; Junior College Students; Junior High School Students; *On Line Systems; *Statistical Analysis
IDENTIFIERS *PLATO IV

ABSTRACT

The objectives of this study were to find ways to use online data for Computer Based Education (CBE) lesson development and evaluation, to find correlation with off-line data, and to find limitations in the use of online data. Lessons in mathematics, biology, accounting, and chemistry were given on PLATO to community college and some high school students. Measures like percent of questions answered correctly on the first try, or ratio of wrong answers to the number of questions, were computed for each student. These measurements were compared among the classes that studied the same lessons. Observation in classes helped in judging whether the data were reasonable. Lessons that showed low performance were changed and the new data were looked at to find a reflection of these changes. Graphic representations of various lesson and student data were generated to help in the comparison. Off-line test scores were correlated with online data and stepwise multiple regression performed. Except in cases of very special lesson programming, it was found that online data can be a useful tool to follow and evaluate lesson development; the few exceptions can be corrected.
(Author/VT)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED165712

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ON-LINE DATA IN CBE

by

Tamar A. Weaver

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

John J. Hirschbuhl

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IR006 275

On-line Data in CBE

by

Tamar A. Weaver

Introduction

The use of CBE allows us to collect data on the student's interaction with lesson materials in as much detail as we want. It is unreasonable to collect full data for a large number of subjects because of space limitations.

Summaries of student interactions in the lessons were collected for a large number of students. The purpose here is to show that measurement of performance and difficulty in lessons can be derived from the data summaries.

The objectives were:

- a. Find ways to use on line data for CBE lesson development and evaluation.
- b. Find correlation with off-line data.
- c. Find limitations in the use of on-line data.

Procedure

Lessons in mathematics, biology, accounting and chemistry, were given on PLATO to community college and some high school students.

Measures like percent of questions answered correctly on the first try or ratio of wrong answers to the number of questions, were computed for each student.

These measurements were compared among the classes that studied the same lessons. Observation in classes helped in judging whether the data were reasonable.

This research was supported in part by National Science Foundation contract US NSF c-723, and by the Advanced Research Projects agency of the Department of Defense under U.S. Army contract DAA c-15-73-c-0077. I want to thank the people that helped in data collection and in editing this paper: Larry Francis, Esther Steinberg, Charles Weaver, R.A. Avner, Lynn Misselt, Keith Bailey, Lou DiBello and Barabara Lederman.

Lessons that showed low performance were changed and the new data were looked at to find a reflection of these changes. Graphic representations of various lesson and student data were generated to help in the comparison.

Off-line test scores were correlated with on line data and stepwise multiple regression performed.

The tools on PLATO IV

The projects involved used the available on-line data machinery on PLATO. These data are summaries of what the student has done in a lesson or parts of it: number of questions presented, number of questions answered correctly on the first try, number of anticipated wrong answers, number of unanticipated wrong answers, number of HELP or branching requests that got the help and number of HELP requests that weren't satisfied. This together with the time it took were collected for each student. A lesson part for which the data is summarized in one summary is called an "area", and therefore the on-line data summaries also called the "area data".

Interpretation of the On Line Data

From the raw data, two main measures were computed. The measure of lesson difficulty was the number of wrong answers per question (no/question) and the measure for smooth and good performance was the percent of questions answered correctly on the first try (okf%).

Comparison of the results in many classes with the instructors' reports and our own classroom observations of lesson difficulty led to a table of standard ranges for these two main measurements. The percent of questions answered correctly on the first try was found to be lower than 70% where there were problems, and higher than 80% when the class was working smoothly without special difficulties.

When an area was hard, the ratio of wrong answers per question was not necessarily high. However, when this ratio was high, a real problem was always present. In practice areas or in trial and error type of instruction, these limits are not relevant.

For each lesson or area (part of a lesson) a normal range of the data was found by taking means for all the classes that used it. These numbers were used later for estimating the match of a lesson to a certain course that used the lesson later on.

Evaluation of Lesson Revisions

The measurements derived from the on-line data in lessons were sensitive to changes in the lessons, and clearly reflected the changes (mainly by the percent of correct answers on first try). The okf% was higher when an improvement was made, and stayed the same when the lesson was not changed. In one case, when a clearly bad change was made, the data reflected decrease in the performance measure as you can see in the next example.

An Example of How the Data Reflects Revisions of a Lesson*

The main instructional part of a math lesson teaching the addition and subtraction of signed numbers was divided into seven areas. Each of the first six areas contained instructions, examples, and exercises dealing with one type of problem. The seventh area provided mixed practice with the previous six problem types. Data for several classes are found in Table 1.

The lesson was first used by over one hundred community college students in Spring 1975. As anticipated by the authors, the difficulty of the fifth and sixth areas (subtracting with negatives) was greater than that of other areas. The final mixed exercise showed more difficulty than expected. In five of the six classes the %okfirst dropped off dramatically for the mixed exercise (the seventh area). You can see that for classes B and C of Table 1, but not for class A.

These data suggested a need to add an additional "help" area before mixing together all six types of problems. Because individual problem types were apparently mastered, the additional part was constructed to teach discrimination between the problem types.

The question was who should have to do this help area. An examination of the histograms of percent ok on first try for each area provided additional insight about what was happening to the students. Figure 1 presents the results of a representative class for the last three areas of the lesson. These and class A of Table 1 suggest that many students apparently did well throughout the lesson whereas a few of the students plummeted rather suddenly on the last mixed exercise. This information, combined with analyses of individual students, led the author to decide that only students

* This example was presented in more detail in a project report "Analysis of Student Data in Computer-based Education" by Tamar a. Weaver and Larry Francis.

Table 1

Class Means in the Signed Number Areas Before Revision

	<u>Area</u>	<u>Min/use</u>	<u>Min/stu</u>	<u>%Okfirst</u>	<u>Errors/ques</u>
<u>Class A</u>	1	2.0	2.3	87	0.40
	2	1.4	1.6	88	0.21
	3	1.7	1.9	88	0.16
	4	1.1	1.2	88	0.22
	5	2.6	2.9	80	0.45
	6	1.6	1.7	78	0.32
	mixed	2.0	2.1	82	0.31
<u>Class B</u>	1	1.4	1.7	85	0.10
	2	1.5	1.8	77	0.35
	3	1.4	1.8	90	0.15
	4	0.9	1.0	88	0.14
	5	2.5	2.7	78	0.35
	6	1.7	1.8	81	0.30
	mixed	2.1	3.2	70	0.48
<u>Class C</u>	1	3.8	5.9	76	0.74
	2	2.3	3.1	78	0.32
	3	2.5	3.2	86	0.16
	4	1.8	2.1	86	0.29
	5	4.7	5.7	70	0.84
	6	4.0	4.6	62	0.56
	mixed	5.1	6.2	56	0.51

who scored lower than 80% on the previous areas would be routed through the additional discrimination exercise (i.e., the "help" area).

A second, simultaneous modification to this lesson was based on instructor requests rather than student data or comments. In order to reduce the time needed for the lesson, the number of examples in each area was reduced from two to one.

These two changes provided the second version of the lesson.

A comparison between the results of the Spring 75, and Fall 75 classes for the first and second versions of the lesson are shown in Figure 2. The classes using the revised version had a uniformly lower performance (as measured by %okfirst).

The reduction in the number of examples was strongly implicated as the reason for the lower performance. These undesirable results suggested that it should be changed back to the original number of examples, and even more in areas that seemed harder.

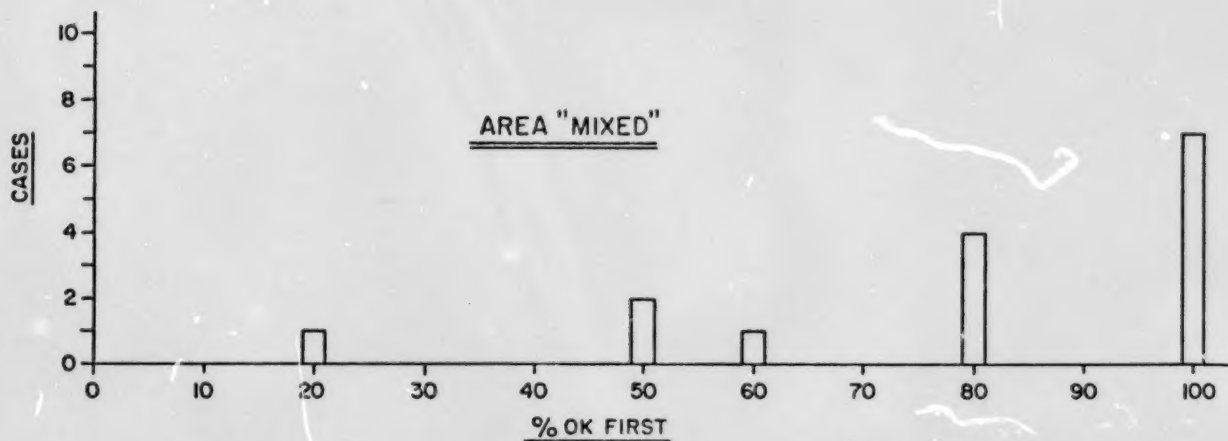
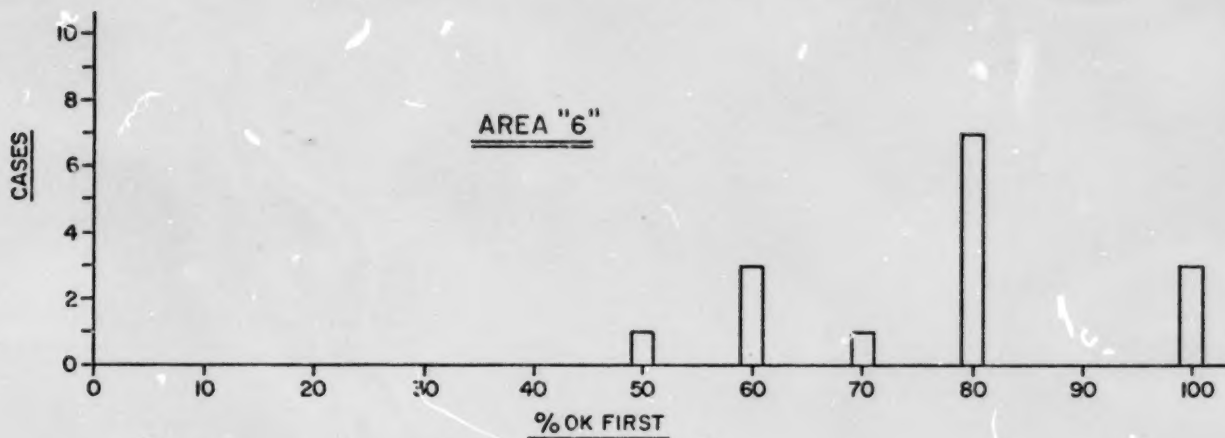
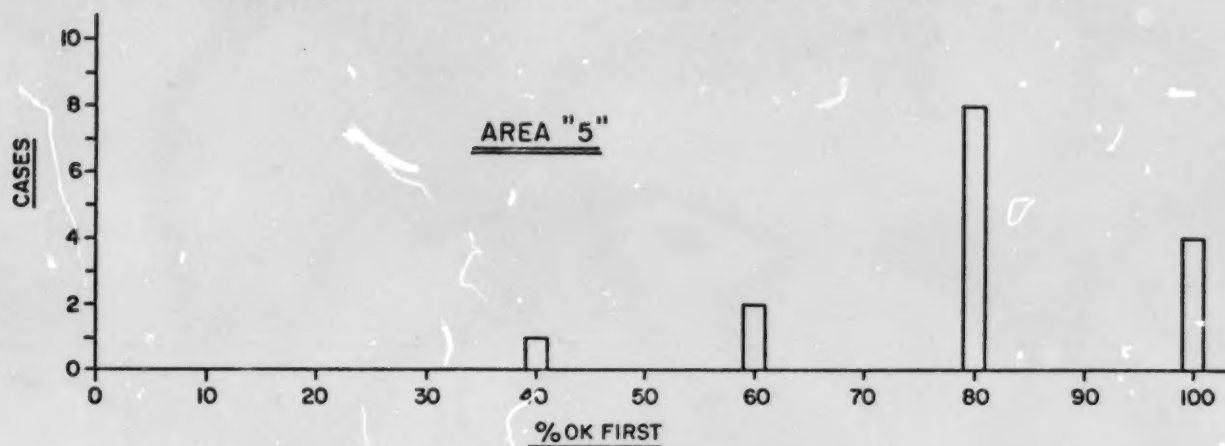
Therefore, one or more examples were added back into all teaching areas except area 4, in which students had scored higher than 80% ok for the revised lesson. Student-by-student analyses indicated the "help" section seemed to be aiding poor students as was intended in spite of the general lower performance.

The data for the third version were collected in Fall 1976. The data from the first two versions were collected from the community college classes, whereas the data from the third version were gathered from junior high school students. Nevertheless, the results of the analysis suggest the lesson behaved similarly for all three groups of students (Figure 2).

The examples added to the third version seem to have raised the %okfirst for all areas except area 4. The fact that the number of examples in area 4 was not changed confirms the validity of the changes to other areas. It is interesting to note that although the %okfirst for this last version is generally lower than that of the first version, the performance on the final mixed test is higher. This is probably due to the "help" area, which was not available to students taking the original form of the lesson. The %okfirst for the help area is lower than all other areas because only students doing poorly in previous sections are routed through

Figure 1

Frequency for Different Okf% in Three Areas Before Revision

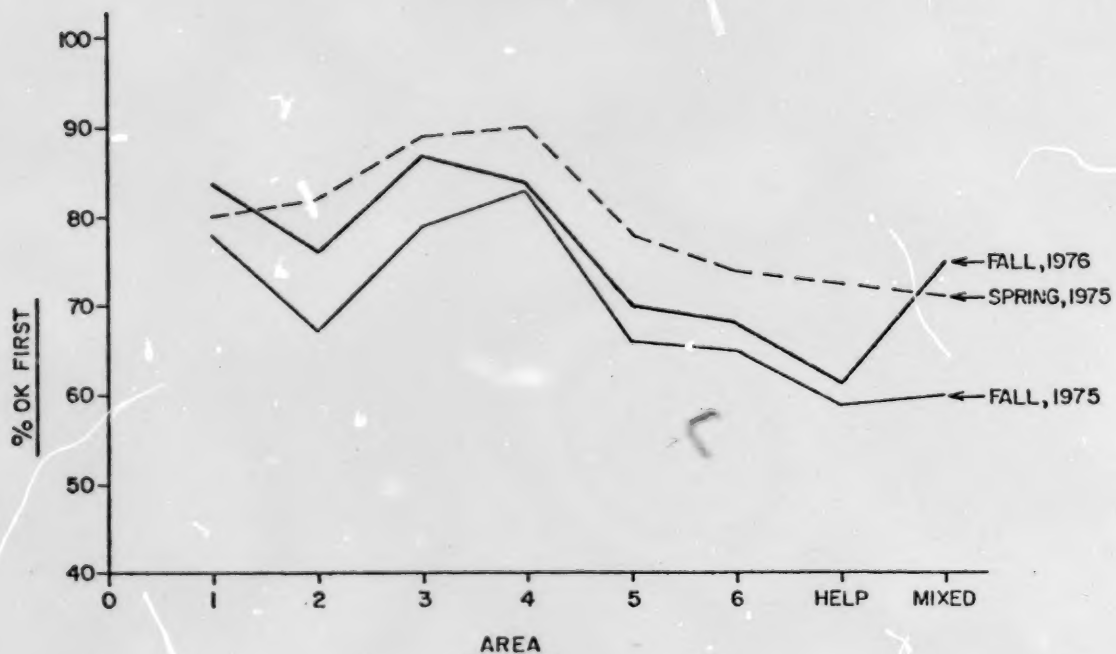


the help area. This extra help and practice in discrimination of problem types tends to keep the poorer students from "falling apart" on the final mixed exercise. The result is that the average score for the whole class, rather than slowly declining through the later areas, makes a sharp upward turn in the final test area. The class average of %okfirst on the final test area is near the average %okfirst of the previous six areas. This is where one would expect it to be if the difficulty with discrimination of problem types had been eliminated since the test samples items from each of the previous areas.

To summarize the example, the few changes in the lesson were very well detected by the student data, and what wasn't changed (area 4) yielded unchanged data.

This example demonstrates how the on-line data reflects very well changes in a lesson and is a good tool for evaluation of such changes.

Figure 2
Comparison between different versions of a lesson



Analysis of Performance Measurements

Any data from an area summary may be transferred to a more powerful set of statistical analysis routines available on the PLATO system. After transfer, the chosen variables can be used in multiple regression routines. Other student data not found in area summaries (like test results) may also be entered in the regression.

Many stepwise multiple regressions were performed to predict test results using on-line data variables as predictors. Three measurements were used: the percent of questions answered correctly on the first try (the variable name in the tables ends with "ok" or %), the wrong answers per question (the variable name ends with "no" or "n") and the time in the area (the variable name ends with "t"). Examples of regression results can be seen in Tables 3, 4 and 5.

The questions asked were which of the area variables will enter in which cases, and which areas will enter for each test.

The expected result was that the "ok" variables would enter with positive contribution to predict test results in a subject relevant to the lesson, the "no" ratio variables would enter with negative contribution to the test results, and the time* would also enter negatively since better students can work through the materials faster. In general, all these things happened.

The way the variables entered in the regression suggests the meaning of these measurements. The okf% (ok on first try) entered considerably in the relevant lessons to predict the test results on the same topic. So the okf% variable may be a good measure for the understanding of specific material.

When the data from irrelevant lesson were checked as predictors to a test, the measurements that were most useful were time and no's per question (both with negative contribution). So these measures may represent general abilities in the field. See Table 2 for examples of such cases.

The few cases where the time or the no per question entered the regression positively, were in practice areas, where the student could practice more if he wanted. The more he practiced, the better he knew the material.

In general, data from lesson that teach material involved in a test, were found to be better predictors of these test results than data from irrelevant lesson. Also, tests on specific skills were better predicted than tests with mixed topics.

* Further time data study was done by K. K. Tatsuoka and M. Tatsuoka.

Table 2

Examples of the Regression Tables for Two Tests

Criterion Variable	20 test2		
Residual Variance	54.767	Degree of Freedom	
Residual Standard Deviation	7.400	Regression	2
Standard Error of Mean	1.795	Error	14
Multiple R	0.799	F-Value	12.322
Squared Multiple R	0.638		
Corrected R for shrinkage	0.614		
Variable Entered	5		
Constant (intercept)	48.47		

Variable-name	B-coef	STD error of B	Beta-coef	STD error of Beta
5 frfbn	-3.230	1.756	-0.322	0.175
12 frfct	-2.102	0.598	-0.615	0.175

Criterion Variable	19 test1		
Residual Variance	25.095	Degree of Freedom	
Residual Standard Deviation	5.0095	Regression	2
Standard Error of Mean	1.2524	Error	13
Multiple R	0.8043	F-Value	11.905
Squared Multiple R	0.6468		
Corrected R for shrinkage	0.6216		
Variable Entered	5		
Constant (intercept)	36.47		

Variable-name	B-coef	STD error of B	Beta-coef	STD error of Beta
5 snbbn	-3.005	1.445	-0.351	0.169
10 dsnet	-3.796	0.983	-0.652	0.169

Note. The variables are described in the previous page.

Examples of regressions for off line tests with the data

In a math 111 class, signed number lessons data were checked as predictors for test results in a few subjects. See test list* at the end of this section.

Test6, which included signed number questions, was predicted very well by okf% in a multiplication of signed number lesson and negatively by no/question in an addition and subtraction of signed numbers lesson ($R^2=.883$; $SR^2=.874$; $F_{2,13}=48.845$; $N=16$). These two measures were good predictors for many of the tests. (SR^2 for shrinkage.)

Test3 covered fractions with signed numbers. It was predicted by the signed numbers areas with $R^2=.777$; $SR^2=.761$; $F_{2,13}=22.677$

In Table 2 there are two examples where the test was not on the subject of the lesson involved: test2, on signed numbers, is predicted by the fraction lesson, and test 1, on set theory, predicted by the signed number data. In both cases the variables that contributed were the time in an area, and the no per question in another area (the time is a significant contributor in both but the no's per arrow is not significant contributor in one of the examples). So the combination of these two variables must be good predictors for general ability in the field rather than for the specific skill. This was repeated in Biology course (Table 5) where a regression for test2 with irrelevant lessons, while test1 (Table 4) is relevant to this same lesson and involved the okf%.

In a chemistry 121 class some math lessons were given on PLATO, along with a lesson about gas laws. In general it turned out that the results of off-line tests containing materials taught by specific PLATO lesson areas were best predicted using area data from those areas. Data from indirectly related lesson areas entered less strongly, and data from irrelevant areas did not appear in the regression at a significant level.

In Test6, which was on gas laws, the most significant predictors were the data from the gas laws lesson. ($R^2=.812$, $F_{2,8}=17.317$; $p=0.001$) The most important variables were the number of no's per arrow in one area, and the % ok on first try in another area of the gas lesson.

* The test data were provided by the Community Colleges of Chicago PLATO project. I want to thank the director of the project, Dr. Chris Dimas, and the instructors, Walter Chamot, Roger Garst, Jane Guagliardo, Glenn Weller, Robert Bator, Jerum Landway, Eddie Ingersol, Lilian Questiaux, Eugene Costabile, Charles Matz, Ruey-Rong Lii.

Test1 did not include questions on gases but rather on mathematics, and its results were not predicted well by the gas lesson data. (For the same variables $R^2=.420$, $F_{2,8}=2.892$; $p>0.05$)

Test2 fell somewhere in between. It was about units of measurement, and was treated indirectly by the PLATO math and gas law lessons. Regressions were made, first using the gas law lesson data and then math lesson data. The results were between those of test 1 and test 2 for the gas laws data: $R^2=.701$, $F_{2,8}=9.4$; for the math lesson data: $R^2=.538$, $F_{1,12}=13.952$, $p=.003$)

Regressions were also made, using data from the PLATO lessons to predict performance on irrelevant tests and none of the data entered into the regressions at a significant level.

Another example where practice or review time contribute positively is in a military mechanics training course with on-line tests. The time in test was predicted with main contribution from no/q in an area of the same lesson, and time in 5 other areas, one of them was a review area, and it entered negatively. ($R^2=.990$; $SR^2=.989$; $F_{6,28}=486$)

An example from a chemistry 201 course test where the time contributes positively is a drill in solutions where the student chooses how many exercises he will do (Tables 3).

The off-line tests in the regression examples

In math 111 class:

- Test1: Set theory
- Test2: Signed Numbers, Graphs, Notations.
- Test3: Fractions
- Test4: Powers, Polynomial expressions
- Test5: Solving Equations
- Test6: Signed Numbers, Fractions, Powers, First degree equations and Set theory

Chemistry 121 tests:

- Test1: Mathematics
- Test2: Units of measurement
- Test3: Atomic structure
- Test4: Chemical Bonds
- Test5: Formulas
- Test6: Phases, Gases
- Test7: solution equations
- Final: All of the above

Table 3
Stepwise regression for test on gas laws, solutions
and the periodic table

$R^2 = .818$, constant = 100.06, $F_{2,5} = 11.255$ $p = .014$

variable	coefficient	S.D. error	t(8)	p
3 drill t	1.320	.481	2.75	.013
15 gas8 t	-1.950	.481	2.75	.025

Table 4
Stepwise regression for biology test1:

$R^2 = .867$, constant = 37.63 $F_{2,12} = 39.007$ $p = 0.000$

variable	coefficient	S.D. error	t(15)	p
a 2 ok	.644	.117	5.504	0.000
a 2 t	-.451	.117	3.855	0.002

Table 5
Stepwise regression for biology test2

$R^2 = .626$, constant = 82.04 $F_{2,12} = 10.004$ $p = 0.003$

variable	coefficient	S.D. error	t(15)	p
a 2 t	-1.121	.250	4.484	0.000
a 35 no	-.761	.250	3.044	0.008

Conclusions

Except with cases of very special lesson programing, it was found that on-line data can be a useful tool to follow and evaluate lesson development. Also the few exceptions can be corrected.

The correlations between student off-line test scores and on-line data measurements show that the time data and the no per question measurements are more measures of aptitude in the field more than knowledge of a certain skill. Percent of correct answer on first try is a measure of the knowledge of the specific skill. This conclusion comes from the last measurement being a predictor to tests that are on the exact same material, while the time and no per question predict tests which are not on the same materials, but from the same field.

Richard Murphy and Lola Appel (1977, pp44-45), of the Educational Testing Service, using the same data and data machinery, claimed that the on line data measures that we provided them with could be used for small, well-designed studies of individual lessons or groups of lessons.

It seems that in each lesson the individual student can only be compared to the class means in the given measures.

It is recommended that the data be used as a warning system rather than a final measure of the student ability.

There is no question that these measurements are a sensitive tool for the evaluation of lesson revisions.

References

- Murphy Richard T. Murphy and Lola Rhea appel.
Evaluatio of The PLATO IV Computer-based Education
System in The Community College.
Educational Testing Service, Princeton, New Jersey, 1977.
- Tatsuoka K. & Siegel, M. Manual to Accompany "area"
summary data and analysis package, Internal report.
Computer-based Education Research Laboratory, University
of Illinois, Urbana, Ill., 1975.
- Tatsuoka, K. & Tatsuoka M. Time-score analysis in criterion-
referenced tests. Progress report to Basic Skill group
of the National Institute of Education for October 1,
1976-April 30, 1977. University of Illinois, Urbana,
Ill., 1977.
- Weaver, T. & Lederman B. A supplemental package for -qarea-
(Chapter 5) in On-line Data Collection compiled by
Dibello, L. V., Community College Math Group Interim
Report, Computer-Based Education Research Laboratory,
University of Illinois, Urbana, Ill., 1975.
- Weaver A. Tamar and Francis Larry, Internal Report for MTC project
Analysis of Student Data in Computer-based Education
University of Illinois, Urbana, Ill., 1975.